DOCUMENT RESUME

ED 435 742 TM 030 394

AUTHOR Miller-Whitehead, Marie
TITLE Bridging the Student Achievement Gap in Science.
PUB DATE 1999-11-19
NOTE 22p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Point Clear, AL, November 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Academic Achievement; Class Size; *Educational Finance; Educational Improvement; Elementary Education; Low Income Groups; Lunch Programs; Poverty; Resource Allocation; *School District Wealth; Science Education; *Science Tests; Sciences; Standardized Tests; *Test Results
IDENTIFIERS *Tennessee

ABSTRACT

The relationship between school system financial and demographic data and student achievement on the science subsection of the Tennessee statewide standardized test was studied to determine if systems that served large numbers of low income students had been as successful as more affluent systems in improving student achievement. System-level data used consisted of Tennessee science scale scores for grades 3 through 8 for 1998, per pupil expenditure, per capita income, percent of students on free and reduced lunch, percent of ethnic group students, and percent of schools in the system meeting state class size guidelines. A hierarchical regression analysis was conducted with variables regressed on student scale scores on the 1998 test. The resultant equation yielded a multiple "R" of 0.82 and an adjusted "R squared" of 0.57, with percent of students on free and reduced lunch accounting for 55% of variability. Across grades 3 through 8, percent of students on free and reduced lunch accounted for the majority of unexplained variance in science scale scores. In a second procedure, for students in grade 5, the same variables regressed on grade 5 science scale score achievement resulted in a formula that yielded a multiple "R" of 0.81 and an adjusted "R squared" of 0.53. At grade 5, percent of students meeting class size standards accounted for the greatest amount of explained variability in science scale scores, with ethnicity marginally significant. Per pupil expenditure had a positive effect on the science scale score achievement for grade 5 students. An appendix maps science scale score trends. (Contains 3 figures, 7 tables, and 26 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 742

Bridging the Student Achievement Gap in Science

Marie Miller-Whitehead, Ph.D.

Paper presented at the annual meeting of the
Mid South Educational Research Association
Point Clear, AL
November 19, 1999

TM030394

BEST COPY AVAILABLE

Education Consultant
PO Box 2882
Muscle Shoals, AL 35662
email: [redacted]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

M Miller - whitehead

2A

Abstract

The purpose of the study was to examine the relationship between school system financial and demographic data and student achievement on the science subsection of the 1998 Tennessee state-wide standardized test to determine if systems which served large numbers of low income students had been as successful as more affluent systems in improving student achievement.

The inquiry used categories from the system level data set from the Tennessee state report card, consisting of the Tennessee science scale scores for grades 3 through 8 for 1998, per pupil expenditure, per capita income, percent of students on free and reduced lunch, percent ethnic, and percent of schools in system meeting state class size guidelines.

Two separate analyses, one for mean scale scores by system across grades 3 through 8, and a second for grade 5, were employed for this study. The first hierarchical regression analysis was conducted with variables regressed on student scale scores on the 1998 Tennessee state test administered to students in grades 3 through 8. The resultant equation yielded a Multiple R of .82 and an adjusted R^2 of .57, with percent of students on free and reduced lunch accounting for 55% of variability. Across grades 3 through 8, percent of students on free and reduced lunch accounted for the majority of explained variation in science scale scores.

In the second procedure, for students in grade 5, the same variables regressed on grade 5 science scale score achievement resulted in a formula which yielded a Multiple R of .81 and an adjusted R^2 of .53. At grade 5, percent of schools meeting class size standards accounted for the greatest amount of explained variability in science scale scores, with ethnicity marginally significant. Per pupil expenditure had a positive effect on the science scale score achievement for grade 5 students.

Paper presented at the annual conference of the MidSouth Educational Research Association, November 1999

In light of the enormous interest in student achievement, school performance indicators, and fiscal accountability which school systems face today, school system administrators must cope with an increasingly complex set of state and federal mandates and initiatives designed to insure that all student populations are provided the best education possible. School and school system accountability issues have also been influenced by funding equity lawsuits in many states, including the State of Tennessee, over the past ten years. The result has been restructuring of school system funding to help provide a more equitable access to excellence in education regardless of the county in which the students attend school.

This issue is made even more complex by accompanying initiatives aimed at increasing site-based decisionmaking at the school level to assure that students receive an educational experience which is reflective not only of federal and state mandates, but of the communities in which they live. Thus, while some school systems may decide that a key objective is to improve student test scores on standardized tests, others systems may target improving student attendance, graduation rates, lowering dropout rates, or other areas of achievement which they have identified as key to assuring that students are provided with every opportunity to become productive members of their communities, to continue their education, or to enter the workforce with the skills desired by 21st century employers. These skills may be very different from county to county, with rural, suburban, and urban employers often not having the same expectations for student graduates as those in neighboring counties not many miles distant. Additionally, systems may prioritize either improvement, meaning that each student will show individual gain or growth in targeted skill areas, or on the other hand, systems may instead set the goal that each student will achieve some predetermined mastery level on criterion referenced tests, thus setting for themselves a cut score which all students are expected to attain. The State of Tennessee has mandated as one component of its legislated accountability system that all students will show improvement and gain in achievement on the state-wide assessments administered to all public school students in Tennessee each April in grades 3 through 8. Thus, while students in Tennessee take both criterion referenced and norm referenced tests in spring of each school year, at the state level systems are held accountable for student gains, or value-added, for each student who attends school in their system.

Given the variety of goals and objectives which school systems may choose to set for themselves, the present study of 1998 science scale score data for Tennessee school systems provides a focus on identifying those factors or variables which continue to effect student achievement on standardized tests. The results of the study are intended to provide school systems with information which will aid them in identifying schools and systems which can be "benchmarked" as setting a standard for achievement and excellence among their peers or school systems in comparable counties with comparable student populations.

This study is a continuation of a longitudinal study of the Tennessee science scale score data beginning with data from 1991, thus now providing eight years of information for use by decisionmakers. Because the purpose of the study was to inform decisionmaking and to identify variables which continue to effect student achievement subsequent to the restructuring of the state of Tennessee's school finance and accountability system under the Educational Improvement Act of

1991, the primary dependent variable for this study was scale scores, rather than the value added gain score computations which is designed to control for these variables. While school systems whose students are at the highest levels of achievement may focus on identifying other areas to improve while holding gains on test scores, most systems would certainly wish to include improvement in student academic achievement as a priority. Since under the value-added system each student is measured against his or her own performance on previous year's tests, the value-added computations are designed to control for student demographic variables such as individual ability, ethnicity, socioeconomic status, and school system characteristics which affect the rate at which students learn. All educators know that all students can learn, but not all will learn the same amount in the same length of time. Using the scale score data rather than the value-added data is an aid in accomplishing four tasks vital to school decisionmakers: 1. How much are students learning compared to other students? 2. What are the factors which are affecting student learning in our system? 3. Are these factors, once identified, related to things which the school system can improve or change so that students achieve to the best of their ability? 4. Are other school systems more successful in helping their students achieve and if so, how?

While the strength of the value-added system lies in its ability to calculate student growth over time, regardless of confounding variables both within and outside of school system and student control (i.e., ability, health, family issues, gender, motivation and interest in school, socioeconomic factors, school climate, teaching, and leadership), scale scores are an additional and vital tool for school and community decisionmakers, providing a quantitative basis for identifying those factors over which teachers, parents, school and system administration, and funding agencies exercise varying amounts of control.

The initial reports in the longitudinal study examined the aggregate school system data to determine differences in student science achievement by grade level and by year statewide for a five year period over grades 2 through 8. Since the test instrument in use at the time, a version of the McGraw Hill CTBS/4, had been normed in the late 1980's, statewide trends and patterns of achievement were more easily identifiable and not affected by changes in the test instrument, norm population, and the like, except for yearly production of the test drawn from items in the McGraw Hill test item pool. The results of these studies indicated that student achievement on the science sub test had shown gradual, but significant improvement over a five year period, which confirmed the findings of earlier studies which had focused on providing evidence of validity and reliability of the TCAP (Tennessee Comprehensive Assessment Program) value-added model as one accountability indicator for Tennessee school systems. Subsequent studies focused on school level and system level data with the inclusion of student and system variables provided in the Tennessee School Report Card.

The school level mean scale scores for all grade levels at each school in each Tennessee system were used to compute system mean scale scores for grades 3 through 8 for a five year period using the MANOVA procedure with 5 levels for year and 6 levels for grade. While the original studies included data for grade 2, due to recent changes in state testing guidelines test score data is currently provided for grades 3 through 8. The data set consisted of 31 variables representing school level data for 138 Tennessee school systems. However, because several Tennessee school systems are not K-8 systems and thus did not have data for all grade levels included in the analysis, 133

systems entered the analysis for the omnibus MANOVA procedure. The decision was made not to replace missing values with the mean as the purpose of the study was to identify significant differences in student achievement from system to system. However, in the preliminary aggregation and examination of results by grade level all systems with grade level test score data entered the analysis for each tested grade level, with frequencies and descriptives for each school system.

This analysis yielded a mean scale score for each school system at each grade level for which student science test scores were available. These data were regressed on the following school system and county indicators made available on the Tennessee State Education Report Card: county per capita income, system per pupil expenditure, percent of students on free and reduced lunch, ethnicity, and percent of schools in system meeting state class size guidelines (Table 1). The 1996 data differed in several key areas from that of 1998, first that the test scale score data no longer included students in grade 2, and second that beginning in spring of 1997 the CTBS/4 was replaced with the McGraw Hill Terra Nova, a new test more recently normed, with subsequently differing range of scale scores from those of the CTBS/4. The previous regression analysis conducted with the 1996 data had yielded an $\underline{R}^2$ of .91 and an adjusted $\underline{R}^2$ of .86 (Miller-Whitehead, 1998, 1999). Of the variables in the analysis, percent free and reduced lunch was by far the most powerful predictor for school system science achievement across grade levels in 1996 ($\underline{r} = -.94$, $p < .001$) with per capita income of the county in which the school system was located having a positive correlation ($\underline{r} = .63$, $p < .05$) to system-wide student performance and per pupil expenditure ($\underline{r} = .46$, $p < .1$) also having a positive correlation to achievement. The 1996 analysis did not, however, include the percent of schools in the system meeting EIA (Educational Improvement Act) class size guidelines. This variable was included in the present study because of initiatives and longitudinal studies which have shown class size to have a significant effect on student achievement (Achilles, 1999; Finn, 1998; Glass & Smith, 1978; Mosteller, 1995; Nye, Achilles, Boyd-Zaharias, Fulton, and Wallenhorst, 1994).

The 1996 study was conducted on system data aggregated across grade levels; no analyses were conducted to determine if there were interaction effects of the variables by grade level. However, for the 1998 test results follow up analyses were conducted for grade level as it was apparent that there was significant variation in rank order student achievement by grade level from system to system. Several systems which entered the analysis for grade 5 did not enter the analysis for grades 3 through 8 as they were not K-8 systems.

Table 1
System Descriptives High and Low Science Scale Scores Across Grades 3 to 8 (n = 21)

System	Science Mean Scale Score 1998	Per Pupil Expenditure 1997-98	% Schools meeting Class Size	% Non-minority	% Free and Reduced 1997-98	County Per Capita Income 1997-98	System Size
A	668.94	$4,351	76.00%	92.7%	6.3%	27,888	15,000 to 25,000
B	667.35	$4,681	100.00%	99.5%	58.0%	12,287	1,000 to 5,000
C	667.32	$4,058	0%	82.4%	29.4%	14,485	less than 1,000
D	665.53	$3,769	100.00%	70.7%	34.2%	15,536	1,000 to 5,000
E	665.28	$4,331	100.00%	91.5%	30.5%	13,740	less than 1,000
F	664.03	$5,335	50.00%	93.7%	21.1%	17,042	1,000 to 5,000
G	664.00	$6,517	71.40%	83.6%	21.4%	19,056	1,000 to 5,000
H	663.32	$4,795	66.70%	92.5%	45.5%	18,161	1,000 to 5,000
I	661.97	$4,273	0%	96.1%	63.3%	15,024	less than 1,000
J	661.35	$4,073	100.00%	95.6%	33.1%	14,890	less than 1,000
P	638.39	$4,487	31.30%	66.0%	40.7%	21,216	35,000 to 50,000
Q	638.20	$3,847	62.50%	99.4%	60.2%	12,247	5,000 to 10,000
R	636.81	$3,774	90.00%	93.7%	48.4%	13,087	1,000 to 5,000
S	636.20	$4,862	52.20%	48.9%	49.8%	18,136	10,000 to 15,000
T	636.04	$4,425	87.50%	98.8%	59.6%	17,845	1,000 to 5,000
U	631.29	$3,777	85.70%	55.9%	52.7%	11,739	1,000 to 5,000
V	630.65	$4,826	100.00%	16.8%	87.5%	15,372	less than 1,000
W	630.48	$5,568	38.20%	50.4%	47.2%	23,804	50,000 to 75,000
X	623.16	$4,871	36.70%	13.3%	67.2%	21,486	75,000 +
Y	619.43	$4,836	100.00%	67.0%	65.2%	16,534	less than 1,000
Z	617.88	$3,988	37.50%	29.5%	82.8%	15,147	1,000 to 5,000
State TN	645.23	$4,391	56.30%	74.0%	41.0%	15,194	890,805

The range of scores for the 10 systems with the highest means was quite small (mean score A = 668.94 and mean score J = 661.35), while for the systems at the low end the range was greater (mean score P = 638.39 and mean score Z = 617.88). The range of scores for the entire sample (n = 21) was 51 scale score points (mean score A = 668.94, mean score Z = 617.88). Systems which had the highest mean scores were more likely to be small systems of less than 5,000 students. However, systems which achieved the lowest mean scores represented both large and small systems. Of the 10 systems which achieved the highest overall science scale scores, only three systems had more than the state average of students on free and reduced lunch (state = 41%, system B = 58%, system H = 45.5%, system I = 63.3%). For systems with the lowest science scale scores, all but one exceeded the state average for percent of students on free and reduced lunch. The school systems in the sample also represented K-12 and K-9 or K-6 grade configurations. The system rank order of science scale score achievement yielded different results when examined for achievement at only

specified grade levels (Table 2). School systems which appeared on both Tables 1 (across all grades) and 2 (grade 5) are designated by the letter assigned on Table 1.

Table 2
Grade 5 Science Scale Score Variables (n = 20)

System	Science Mean Scale Score 1998 Grade 5	Per Pupil Expenditure 1997-98	% Schools meeting Class Size	%ethnic	% Free and Reduced 1997-98	County Per Capita Income 1997-98	System Size
K	684.25	$4,225	100.00%	98.6%	66%	13,908	1,000 to 5,000
L	675.60	$5,269	100.00%	98.0%	37%	19,056	1,000 to 5,000
D	673.90	$3,769	100.00%	70.7%	34%	15,536	1,000 to 5,000
A	673.65	$4,351	76.00%	92.7%	06%	27,888	15,000 to 25,000
F	670.38	$5,335	50.00%	93.7%	21%	17,042	1,000 to 5,000
G	665.60	$6,517	71.40%	83.6%	21%	19,056	1,000 to 5,000
M	665.40	$5,522	100.00%	92.7%	18%	27,888	1,000 to 5,000
N	664.25	$5,918	85.70%	98.5%	30%	14,591	1,000 to 5,000
O	664.00	$4,331	100.00%	91.5%	31%	13,740	less than 1,000
AA	663.93	$4,889	100.00%	94.9%	37%	11,300	1,000 to 5,000
AB	637.80	$3,665	50.00%	96.1%	34%	16,558	less than 1,000
AC	637.70	$3,651	0%	88.8%	50%	15,396	less than 1,000
AD	637.37	$3,767	100.00%	99.1%	38%	14,171	1,000 to 5,000
W	636.50	$5,568	38.20%	50.4%	47%	23,804	50,000 to 75,000
AE	634.30	$3,850	66.70%	85.9%	34%	12,500	1,000 to 5,000
Q	633.48	$3,847	62.50%	99.4%	60%	12,247	10,000 to 15,000
AF	632.00	$4,313	100.00%	44.2%	57%	16,558	1,000 to 5,000
V	631.00	$4,826	100.00%	16.8%	88%	15,372	1,000 to 5,000
X	630.78	$4,871	36.70%	13.3%	67%	21,486	75,000 +
Z	617.90	$3,988	37.50%	29.5%	83%	15,147	1,000 to 5,000
State	650.87	$4,391	56.30%	74.0%	41.0%	15,194	890,805

When aggregated at the grade level by system, three of the top 10 performing systems appeared among those whose students achieved highest mean scores for grade 5 science and for the system as a whole (system A, D, and F) while for systems with lower science achievement mean scores systems (system Q, W, V, X, and Z) five of the systems were among the 10 lowest performing systems for grade 5 and also for the system as a whole. There was a greater range of scores for science achievement at grade 5 (mean scale score for system K= 684.25, mean scale score for system Z = 617.90, a range of 66 scale score points) than for science achievement over all grades 3 through 8, with a range of 51 scale score points. While smaller systems once again achieved the highest science scores at grade 5, systems with lower mean science achievement scores included both small and large to very large school systems.

Table 3
Descriptives by Grade Level for 1998 Tennessee Science Scale Scores, All Systems

Grade	N	Range	Minimum	Maximum	Mean	S.E.	Std	Variance
3	137	73.34	573	647	617	.96	11.2	126.07
4	137	61.66	599	660	634	.86	10.1	102.25
5	137	66.35	618	684	651	.80	9.3	86.62
6	136	55.83	625	681	660	.79	9.2	85.42
7	134	57.90	636	694	671	.83	9.6	92.34
8	134	65.95	644	710	685	.91	10.6	111.67

Science scale scores for 1998 and demographic data were examined to determine the relationship of the following variables on students' 1998 science achievement: percent of students on free and reduced lunch, percent of schools meeting state class size guidelines, county per capita income, system per pupil expenditure, and percent of non-minority (white) students. The designation of percent of minority students has become somewhat problematic as in many school systems minority students comprise the majority of the student population. Hierarchical regression was determined to provide an approach which would account for both direct and indirect effects of the variables in the analysis and to allow a more parsimonious selection of variables. The selection of variables in hierarchical regression can be quite complex (Cook & Campbell, 1979; Bryk & Raudenbush, 1992; Mohr, 1995; Yates, 1996) and while many variables were available for inclusion in this study, the decision was made to limit these to those which the literature supported as having the greatest relationship to student achievement, in particular class size, poverty status, ethnicity, county per capita income, and per pupil expenditure.

Thus, not only would poverty be expected to be negatively related to student test scores, there might also be an indirect relation of ethnicity and poverty to test scores (Anthony & Jacobson, 1992). County per capita income, while expected to have a negative relation to percent of students at poverty level, did not play as significant a role in per pupil expenditure as other financial indicators not included as variables in this study.

Table 4
Pearson's Product Moment Correlations, Science Scale Scores Across Grades 3 through 8

	sscore	% clssize	% ethnic	% f/r	percapita$	perpupil$
sscore	1.00	.03	.69	-.73	-.01	.06
% clssize	.03	1.00	.15	-.02	-.24	-.09
% ethnic	.69	.15	1.00	-.56	-.19	-.13
% f/r	-.73	-.02	-.56	1.00	-.36	-.18
percapita$	-.01	-.24	-.19	-.36	1.00	.46
perpupil$	.06	-.09	-.13	-.18	.46	1.00

Proposed Path Analysis Model:

ss98 β_s
%FR β_f
%ethnic β_w
clssize β_c
perpup$ β_p
percap$ β_i

Zero-Order Betas:

$\beta_{fs} = -54.70$
$\beta_{ws} = 21.70$
$\beta_{cs} = -.044$
$\beta_{ps} = .002$
$\beta_{is} = -.001$

$\beta_{fw} = -.56$
$\beta_{ip} = .46$
$\beta_{fw*s} = 30.61$
$\beta_{ip*s} = .01$

Figure 1

A hierarchical regression entering all six variables in the first equation resulted in a Multiple R of .82 and an adjusted R^2 of .57. After adjusting for sample size (n=21), these six variables accounted for 57% of variability in 1998 science scale scores across grades 3 through 8. Percent of students on free and reduced lunch had the greatest effect on science scale scores, accounting for 55% of variability. The science scale score across grades decreased by a factor of 55 times percent of students on free and reduced lunch. For ethnicity, the science scale score increased by a factor of 21.7 times percent of caucasian or non-minority students. Science scale scores across grades 3 through 8 decreased by a factor of .04 times percent of system meeting class size standards. For per capita income, science scale scores decreased by a factor of .0010 times county percapita income. For system per pupil expenditure, science scale scores increased by a factor of .0024 times perpupil expenditure. Of the variables, only percent of students on free and reduced lunch was statistically significant. However, as the number of cases per variable was less than that recommended for regression, the analysis may have underestimated the actual significance. There continues to be considerable debate over the issues of sample size and statistical significance (Kaufman, 1998; McLean and Ernest, 1998; Levin, 1998; Thompson, 1998; Chow, 1996; Cohen, 1987), however given the large effect size yielded by the variables in the regression power for the cumulative R^2 at $\alpha = .05$ was determined to be .89 for a sample size of 20 (Stevens, 1990).

Figure 2

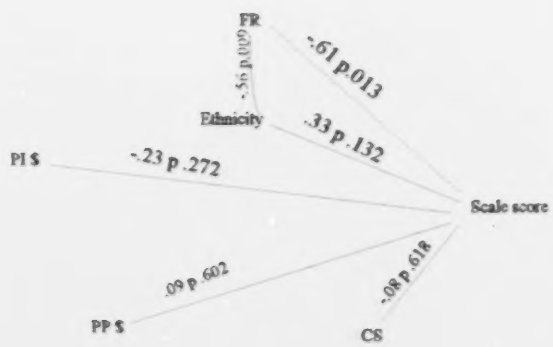

Table 5
Summary of Hierarchical Regression Analysis for Variables Predicting 1998 Science Scale Score Mean Across Grades 3 - 8

Variable	β	Beta	Sig T
Step 1			
Percent free and reduced lunch	-54.7	- .61	.013
Percent ethnic	21.7	.33	.132
Class size	- .044	- .08	.618
Per pupil $	.0024	.09	.602
County per capita $	- .0010	- .23	.272
Step 2			
Percent ethnic	- .410	- .56	.009
Step 3			
Per pupil $	.077	.46	.036

Because of the negative correlation between percent of white students and percent of students on free and reduced meal status, to examine indirect effect of ethnicity on student science achievement, percent of students on free and reduced meal status was regressed on percent of non-minority students. The resultant equation yielded a Multiple R of .56 and an adjusted R^2 of .27, with the model accounting for 27% of the variability in percent of students on free and reduced meal status. Percent of students on free and reduced lunch increased by a factor of .41 for every percent increase in minority students. Thus, while percent ethnicity did not have a significant direct effect on test scores (.33 p = .132), it did have a significant indirect effect (.56 p = .009) and thus actually accounted for more variability in student test scores (.61 x .56 = .34) than that explained by its direct effect.

Since county per capita income might be presumed to have an indirect effect on perpupil expenditure, system per pupil expenditure was regressed on county per capita income to determine its indirect effect on student achievement. The formula resulted in an adjusted R^2 of .17, thus 17% of system per pupil expenditure was explained by county per capita income, with system per pupil expenditure increasing by a factor of .077 times county per capita income.

The model resulted in the following equations for each of the dependent variables:

SS98:

$\gamma' = 666.54 - 55(\%f/r) + 21.7(\%white) - .044(clssize) - .001(percap\$) + .0024(perpup\$)$

F/R

$\gamma' = .778 - .41(\%white)$

perpup$:

$\gamma' = 3247.74 + .077\ (percap\$)$

Actual values for several systems were substituted in the equation (before controlling for indirect effects) to compare actual with predicted results for each of the variables in the analysis:

System A:

SS98:

$\gamma' = 666.54 - 55(.063) + 21.7(.927) - .044(.76) - .001(27,888) + .0024(4,351) = 665.71$

actual = 668.94

System B:

SS98:

$\gamma' = 666.54 - 55(.58) + 21.7(.995) - .044(1.00) - .001(12,287) + .0024(4,681) = 652.10$

actual = 667.35

System H:

SS98:

$\gamma' = 666.54 - 55(.455) + 21.7(.925) - .044(.667) - .001(18,161) + .0024(4,795) = 654.91$

actual = 663.32

System I:

SS98:

$\gamma' = 666.54 - 55(.633) + 21.7(.961) - .044(0) - .001(15,024) + .0024(4,273) = 647.81$

actual = 661.97

System W:

SS98:

$\gamma' = 666.54 - 55(.472) + 21.7(.504) - .044(.382) - .001(23,804) + .0024(5,568) = 641.06$

actual = 630.48

System Z:

SS98:

$\gamma' = 666.54 - 55(.828) + 21.7(.295) - .044(.375) - .001(15,147) + .0024(3,988) = 621.81$
actual = 617.88

Systems A, B, H, and I student science scale score achievement exceeded predicted achievement, with system A and I having both higher student achievement and lower per pupil expenditure than predicted by the model. System B and H exceeded both predicted student achievement and per pupil expenditure, while system Z (a small system) had lower than predicted student achievement across grades 3 through 8 and lower system per pupil expenditure. The model was able to predict science scale score achievement accurately to within less than one standard deviation (18.32 points) and in several cases within 5 scale score points. Although not statistically significant, percent of schools in system within state class size guidelines had a slight negative effect on student science achievement over grades 3 through 8. Because the literature strongly suggests that smaller class size has a positive effect on student achievement in the earlier grades (Achilles, 1999; Achilles & Price, 1999; Finn, 1998; Glass & Smith, 1978; Hanushek, 1998; Nye et al., 1994; Mosteller, 1995), a followup examination of the variables was conducted with grade 5 mean science scale score achievement as the dependent variable to determine if the class size variable had a different effect by grade level.

Table 6
Pearson's Product Moment Correlations for Grade 5 Science

	ss98.5	% clssize	% ethnic	% f/r	percapita$	perpupil$
ss98.5	1.00	.49	.57	- .61	.19	.36
% clssize	.49	1.00	.19	- .17	- .10	.13
% ethnic	.57	.19	1.00	- .66	- .20	- .09
% f/r	- .61	- .17	- .66	1.00	- .61	- .29
percapita$	.19	- .10	- .20	- .37	1.00	.41
perpupil$	.36	.13	- .09	- .29	.41	1.00

Proposed Path Analysis Model Grade 5 Science Scale Scores

Subscripts:

ss98.5	β_s
%FR	β_f
%ethnic	β_w
clssize	β_c
perpup$	β_p
percap$	β_i

Zero-Order Betas:

β_{fs} = - .80
β_{ws} = 40.10
β_{cs} = .25
β_{ps} = .006
β_{is} = .001

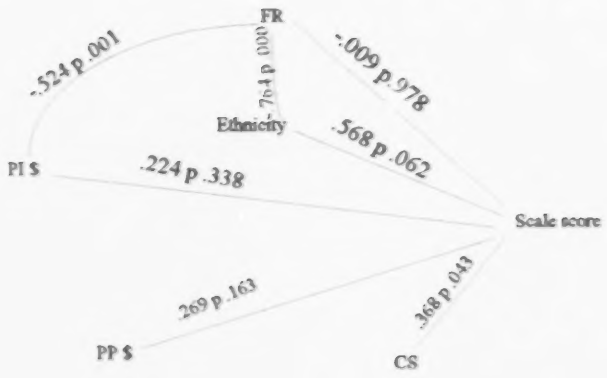

Figure 3

Table 7
Summary of Hierarchical Regression for Variables Predicting Grade 5 Science Scale Score Achievement

Variable	β	Beta	Sig T
Step 1			
Percent free and reduced lunch	- .802	- .009	.978
Percent ethnic	40.10	.568	.062
Class size	.247	.368	.043
Per pupil $	.0064	.269	.163
County per capita $	.00094	.224	.338
Step 2			
Percent ethnic	- .579	- .764	.000
County per capita $	- .000024	- .524	.001

A hierarchical regression entering the six variables in the first equation yielded a Multiple R of .81 and an adjusted R^2 of .53. After adjusting for sample size (n=20), the five variables accounted for 53% of variability in the 1998 science scale score achievement of grade 5 students. Of the variables in the analysis, only class size was statistically significant, accounting for 37% of the total variability in grade 5 science achievement. Science scale scores increased .25 point for each percent increase in schools which met state class size guidelines. A stepwise regression procedure with the same variables resulted in the removal of all but two variables (percent of students on free and reduced lunch and class size) which yielded an adjusted R^2 of .47, less than the variability accounted for by the entry method. For a sample size of 20 at an alpha of .05, power for the analysis was determined to be .82.

To determine the indirect effect of the variables on science achievement, percent of students on free and reduced lunch was regressed on percent white students (non-minority) and county per capita income and the resultant equation yielded a Multiple R of .84 and an adjusted R^2 of .66. While the indirect effect of variables on class size was of interest, the variables in the present analysis provided insignificant R^2 changes when regressed on class size and thus were not included in the model.

The model yielded the following equations for the variables which entered each step of the analysis:

SS98.5:

$\gamma' = 557.39 - .80(\%f/r) + .25(clssize) + 40(\%ethnic) + .006(perpup\$) + .0009(percap\$)$

F/R

$\gamma' = 1.27 - .58(\%ethnic) - .00002(percap\$)$

Actual values were substituted in the equation (before controlling for indirect effects) to compare predicted with actual values with the following results:

System K:

$\gamma' = 557.39 - .80(66) + .25(100) + 40(.986) + .006(4,225) + .0009(13,908) = 606.90$
actual = 684.25

System D:

$\gamma' = 557.39 - .80(34) + .25(100) + 40(.707) + .006(3,769) + .0009(15,536) = 674.46$
actual = 673.90

System A:

$\gamma' = 557.39 - .80(6) + .25(76) + 40(.927) + .006(4,351) + .0009(27,888) = 632.40$
actual = 673.65

System M:

$\gamma' = 557.39 - .80(18) + .25(100) + 40(.927) + .006(5,522) + .0009(27,888) = 663.30$
actual = 665.40

System W:

$\gamma' = 557.39 - .80(47) + .25(38.2) + 40(.504) + .006(5,568) + .0009(23,804) = 604.33$
actual = 636.50

System X:

$\gamma' = 557.39 - .80(67) + .25(36.7) + 40(.133) + .006(4,871) + .0009(21,486) = 566.86$
actual = 630.78

System Z:

$\gamma' = 557.39 - .80(83) + .25(37.5) + 40(.295) + .006(3.988) + .0009(15,147) = 549.73$
actual = 617.90

The systems for which variables for percent ethnic and percent free and reduced lunch

were nearer the mean (i.e., system D) exhibited much closer model fit before controlling for indirect effects than those systems which varied most from the mean (System K, A, W, Z) for ethnicity and poverty. Before controlling for indirect effects, systems W, X, and Z had student achievement results considerably higher than predicted by the model, based on values for the variables used in the equation. The underestimation of the model may be indicative of a curvilinear relation for some of the variables and transformation of these may have resulted in better model fit; however, difficulty of interpretation of results would negate the usefulness of this approach.

Discussion

Tennessee provides each school system with value-added scores based on computation of student performance over a three-year period, the results of which provide a value-added gain score of teacher, school, and school system performance. This procedure thus controls for the confounding variables of student ability, ethnicity, socioeconomic status, school efficacy, community involvement and support, class size, and the myriad other confounding variables which affect student achievement. The present study provides a one year single point in time analysis with the purpose of determining at which grade levels and for which school systems variables continue to play a significant role in student achievement on the science subtest of the statewide test. Therefore, results may differ from subtest to subtest and reading, language arts, social studies or math scale scores might not have resulted in the same equation for the variables. While the rationale for using scale scores as a measure is to provide a method for comparison of scores on the same subtest, this does not apply to comparisons across subtests. The variable for class size may well have a different effect for reading and language arts depending upon grade level than the science results.

Thus, while across grade levels the results of this study indicate that smaller class size was not significant in increasing student science achievement, at grade 5 school systems which were at state recommended class size realized an increase of up to 25 scale score points on the science subtest. This is an indication that class size continues to have a positive effect on student test achievement through grade 5. It is also important to note that science scale score achievement of students in grades 6, 7, and 8 (the across grade data) only accounts for percent of school systems meeting class size requirements in 1998 and thus does not take into account the possible effect on student achievement of class size or other variables in prior years (Sanders & Rivers, 1996). In addition, while maximum class size for grades K-3 is 25, and an average of 20, this number increases to a maximum of 30 for grades 4-6 and 35 for students in grades 7 and 8 (TCA 49-1-104). Though class size recommendations are based on many more factors than can be measured by a standardized test (French, 1993), meeting or exceeding state class size guidelines is certainly indicated for those schools concerned with improving student achievement, as this is one variable which is well within the control of a school system and which can help to offset disadvantages due to student poverty and ethnicity so that all students have an opportunity to be successful in school.

For grade 5 students class size accounted for 37% of the explained variability in science scale

scores, with ethnicity marginally significant. The percent of students on free and reduced lunch had little direct effect on scale scores at grade 5. This finding is indicative that Tennessee school systems have implemented effective programs for improving student science achievement of low income students in grade 5. Thus, the value-added gain scores for a three year period provide valuable information on the cumulative and residual effects of variables on student achievement for each system. Across grades 3 through 8, percent of students on free and reduced lunch accounted for 61% of variation in student science achievement, with ethnic composition of the school system having a direct effect of 33% as well as an indirect effect by meal status. Although across grades 3 through 8 per pupil expenditure had little direct effect on student science achievement (9%), per pupil expenditure accounted for 27% of variability at grade 5.

From the perspective of the effective schools movement (Lezotte, 1989; 1993) the findings indicate that while many Tennessee systems have science scale score achievement higher than expected based on system demographics, others (those with extreme values on the variables) should examine a variety of alternatives to improve opportunities to learn so that their student achievement is more evenly distributed across student groups, i.e., by ethnicity, gender, and socioeconomic status.

References

Achilles, C. M. (1999). *Let's Put Kids First, Finally: Getting Class Size Right*. Thousand Oaks, CA: Corwin Press.

Achilles, C. M., and Price, W. J. (1999). Can your district afford smaller classes in grades K-3? Can smaller classes be cost effective? *School Business Affairs, 65*(1), 10-16.

Anthony, P. and Jacobson, S. (1992). *Helping At-Risk Students: What Are the Educational and Financial Costs?* Newbury Park, CA: Corwin Press.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models.* Newbury Park, CA: Sage Publications.

Chow, S. L. (1996). *Statistical Significance: Rationale, Validity, and Utility*. London: Sage Publications.

Cohen, J. (1987). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.

Finn, J. D. (1998). *Class Size and Students at Risk: What Is Known? What Is Next?* Washington, D. C: U.S. Department of Education. Office of Educational Research and Improvement. (AR 98 7104)

French, N. K. (1993). Elementary teacher stress and class size. *Journal of Research and Development in Education, 26*(2), 66-73.

Glass, G. V., and Smith, M. L. (1978). *Meta-analysis of Research on the Relationship of Class Size and Achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

Hanushek, E. A. (1998). *The Evidence on Class Size*. Rochester, NY: The University of Rochester, W. Allen Wallis Institute.

Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools, 5*(2), 1-3.

Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? *Research in the Schools, 5*(2), 43-54.

Lezotte, L. W. (1989). Base school improvement on what we know about effective schools. *American School Board Journal, 176*(8), 18-20.

Lezotte, L. W. (1993). Creating effective schools today and tomorrow. *Journal for Quality and Participation, 16*(1), 22-30.

McLean, J. E., and Earnest, J. M. (1998). The role of statistical significance testing in educational research. *Research in the Schools, 5*(2), 15-22.

Miller-Whitehead, M. (1998). An analysis of science scale scores for grades 2-8 in Tennessee for 1990 to1994 (ERIC Document Reproduction Service No. ED 414 160)

Miller-Whitehead, M. (1998). A longitudinal analysis of science scale scores grades 2-8 in Tennessee for 1992-1996 (ERIC Document Reproduction Service No. ED 414 170)

Miller-Whitehead, M. (1999). Tennessee TCAP science scale scores 1990-1997: Implications for continuous improvement and educational reform (ERIC Document Reproduction Service No. ED 416 070)

Mohr, L. B. (1995). *Impact Analysis for Program Evaluation.* Thousand Oaks, CA: Sage Publications.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113-127.

Nye, B. A., Achilles, C. M., Boyd-Zaharias, J., Fulton, B. D., and Wallenhorst, M. (1994). Small is far better. *Research in the Schools, 1*(1), 9-20.

Sanders, W. L., and Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville: The University of Tennessee Value-Added Research and Assessment Center.

Steven, J. (1990). *Intermediate Statistics: A Modern Approach.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools, 5*(2), 33-38.

Yates, B. T. (1996). *Analyzing Costs, Procedures, Processes, and Outcomes in Human Services.* Thousand Oaks, CA: Sage Publications.

APPENDIX

Five Years Trends of System Means for High and Low Achieving Tennessee Systems